

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 Perkara Pertanahan

Murad 1991, menyebutkan bahwa timbulnya sengketa hukum atas tanah bermula dari pengaduan satu pihak, pengaduan tersebut berisi tentang keberatan atau pun tuntutan hak atas tanah baik terhadap status maupun prioritas kepemilikannya dengan harapan dapat memperoleh penyelesaian yang sesuai dengan peraturan yang berlaku.

Peraturan Badan Pertanahan Nasional No. 3 Tahun 2011 menyebutkan bahwa Perkara pertanahan merupakan perselisihan pertanahan antara seseorang, kelompok, golongan organisasi, badan hukum atau lembaga yang mempunyai kecenderungan atau sudah berdampak luas secara sosio politis.

2.1.1 Penyebab Perkara Pertanahan

Menurut Undang-Undang PTUN No. 5 Tahun 1996, penyebab umum timbulnya perkara pertanahan dibagi atas:

1. Tumpang tindih peraturan.
2. Tumpang tindih peradilan.
3. Tumpang tindih penggunaan tanah.
4. Nilai ekonomis tanah tinggi.
5. Kesadaran masyarakat meningkat.
6. Tanah tetap, penduduk bertambah.
7. Kemiskinan.

2.1.2 Tipologi Kasus Pertanahan

Tipologi kasus pertanahan yang terjadi merupakan jenis sengketa, konflik atau perkara pertanahan yang disampaikan oleh Badan Pertanahan Nasional, secara garis besar dapat dikelompokkan menjadi beberapa kelompok, antara lain sebagai berikut:



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Penguasaan tanah tanpa hak adalah perbedaan pendapat mengenai status penguasaan tanah yang belum diberi hak (tanah negara).
2. Sengketa batas, yaitu perbedaan pendapat atas suatu bidang tanah dimana satu pihak telah ditetapkan oleh Badan Pertanahan Nasional.
3. sengketa waris yaitu perbedaan pendapat mengenai status penguasaan atas suatu tanah yang berasal dari warisan.
4. Jual berkali-kali yaitu perbedaan pendapat mengenai status penguasaan tanah yang diperoleh dari jual beli lebih dari 1 orang
5. Sertifikat ganda, yaitu perbedaan pendapat mengenai status penguasaan tanah yang memiliki sertifikat hak atas tanah lebih dari 1 orang.
6. Sertifikat pengganti yaitu perbedaan pendapat mengenai status penguasaan tanah yang memiliki sertifikat hak atas pengganti.
7. Akta jual beli palsu yaitu perbedaan pendapat akan suatu bidang tanah tertentu karena adanya akta jual beli palsu.
8. Kekeliruan penunjukan batas yaitu perbedaan pendapat atas suatu bidang tanah yang letak, luas dan bata tanah diakui satu pihak yang sudah ditetapkan oleh Badan Pertanahan Nasional Republik Indonesia berdasarkan penunjukan batas tanah yang salah.
9. Tumpang tindih yaitu perbedaan pendapat akan suatu bidang tanah yang diakui satu pihak tertentu karena terdapat tumpang tindih batas kepemilikannya.
10. Putusan pengadilan yaitu perbedaan pendapat dari kepentingan mengenai suatu putusan badan peradilan yang berkaitan dengan subyek dan obyek hak atas tanah.

2.2 Peradilan Tata Usaha Negara (PTUN)

Dalam Pasal 1 Undang-undang Nomor 5 Tahun 1996, diuraikan tentang pengertian-pengertian yang berkaitan dengan Peradilan Tata Usaha Negara, sebagai berikut:

1. Tata Usaha Negara adalah administrasi negara yang melaksanakan fungsi untuk menyelenggarakan urusan pemerintahan, baik di pusat maupun di daerah.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Badan Tata Usaha Negara adalah Badan yang melaksanakan urusan pemerintahan berdasarkan undang-undangan yang berlaku.
3. Keputusan Tata Usaha Negara (KTUN) adalah penetapan tertulis yang dikeluarkan oleh Badan Tata Usaha Negara yang berisi tindakan hukum tata usaha negara yang berdasarkan undang-undangan yang berlaku.
4. Sengketa Tata Usaha Negara adalah sengketa yang timbul dalam bidang Tata Usaha Negara antara orang atau badan hukum perdata dengan Badan Tata Usaha Negara, baik yang ada di pusat maupun yang ada di daerah.
5. Gugatan Tata Usaha Negara adalah permohonan yang berisi tuntutan terhadap Badan atau Pejabat Tata Usaha Negara dan diajukan ke pengadilan untuk mendapatkan keputusan.
6. Tergugat adalah Badan Tata Usaha Negara yang mengeluarkan keputusan berdasarkan wewenang yang ada padanya, yang digugat oleh orang atau badan hukum perdata.
7. Dalam Pasal 53 ayat (1) Undang-undang Nomor 9 Tahun 2004 menyebutkan bahwa penggugat adalah setiap orang atau Badan Hukum Perdata yang merasa kepentingannya dirugikan akibat dikeluarkannya Keputusan Tata Usaha Negara.

Proses pengajuan gugatan ke PTUN ada beberapa hal yang harus dipenuhi oleh penggugat, diantaranya :

1. Identitas penggugat. Identitas yang dimaksud adalah seperti nama, alamat, kewarganegaraan, pekerjaan.
2. Identitas dari tergugat. Identitas tergugat berupa nama, dan alamat dari tergugat.
3. Melampirkan data akta yang dimiliki. Akta yang dimaksud merupakan akta hak tanah yang dimiliki oleh penggugat. Adapun hak-hak tersebut diatur dalam pasal 16 Undang-Undang Peraturan Agraria, yaitu :
 - a. Hak Milik
 - b. Hak Guna Usaha



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- c. Hak Guna Bangunan
- d. Hak Pakai
- e. Hak Membuka Tanah

4. Menjelaskan objek gugatan yang menjadi perihal penggugat mengajukan gugatan. Objek gugatan merupakan hal atau apa yang digugat oleh penggugat. Dalam hal ini merupakan KTUN yang dikeluarkan oleh pejabat TUN. Menurut pasal 106 ayat 1 Peraturan Menteri Agraria No. 9 Tahun 1999, yang menjadi objek gugatan adalah :

- a. Surat Keputusan (SK) yang dikeluarkan terdapat kesalahan prosedur.
- b. Surat Keputusan (SK) yang dikeluarkan terdapat kesalahan penerapan Undang-Undang.
- c. Surat Keputusan (SK) yang dikeluarkan terdapat kesalahan objek hak.
- d. Surat Keputusan (SK) yang dikeluarkan terdapat tumpang tindih hak.

5. Melampirkan alasan gugatan yang mendasari gugatan penggugat. Ada beberapa alasan yang mendasari adanya gugatan atas KTUN yang dikeluarkan oleh PTUN. Alasan-alasan ini diatur dalam pasal 53 ayat 2 Undang-Undang No. 9 Tahun 2004, yaitu sebagai berikut :

- a. KTUN tersebut bertentangan dengan undang-undang yang berlaku.
- b. KTUN tersebut tidak prosedural dan tidak substansial.
- c. KTUN tersebut dikeluarkan oleh pihak yang tidak berwenang.
- d. KTUN tersebut bertentangan dengan asas-asas umum pemerintahan yang baik.
- e. KTUN tersebut dikeluarkan tidak melalui proses inventarisasi dan pemeriksaan terhadap fakta-fakta yang relevan.
- f. KTUN tersebut bersifat konkrit, individual dan final serta menimbulkan akibat hukum.
- g. KTUN tersebut tidak memerlukan persetujuan atasan atau instansi lain.
- h. KTUN tersebut merupakan perbuatan tergugat yang sewenang-wenang dan merugikan penggugat.
- i. KTUN tersebut menyebabkan hak-hak penggugat sebagai pemilik tanah terabaikan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

j. Gugatan yang diajukan oleh penggugat sesuai dengan waktu yang dibenarkan.

6. Melampirkan permohonan penggugat yang diharapkan agar dikabulkan oleh majelis hakim. Permohonan ini diatur dalam pasal 53 ayat 2 Undang-Undang No. 9 Tahun 2004, yaitu sebagai berikut :

- a. Pembatalan atau menyatakan tidak sah KTUN yang dikeluarkan.
- b. Ganti rugi.
- c. Rehabilitasi.

7. Membayar biaya pendaftaran kasus ke administrasi PTUN.

2.2.1 Kewenangan dalam Peradilan Tata Usaha Negara

Dalam peradilan Tata Usaha Negara ada beberapa hal yang menjadi kewenangan dalam Peradilan, antara lain sebagai berikut:

1. Keputusan Tata Usaha Negara (KTUN). Keputusan Tata Usaha Negara adalah penetapan tertulis yang dikeluarkan oleh badan Tata Usaha Negara.
2. Dikeluarkan oleh Badan TUN. Sebagai suatu Keputusan TUN, Penetapan tertulis itu juga merupakan salah satu instrumen yuridis pemerintahan yang dikeluarkan oleh Badan atau Pejabat TUN dalam rangka pelaksanaan suatu bidang urusan pemerintahan.
3. Berisi tindakan Hukum TUN. Untuk dapat dianggap suatu Penetapan Tertulis, maka tindakan Badan TUN itu harus merupakan suatu tindakan hukum, artinya dimaksudkan untuk menimbulkan suatu akibat hukum TUN.
4. Pasal 1 angka 2 Undang-Undang No 5 Tahun 1986 menyebutkan bahwa semua peraturan yang bersifat mengikat secara umum, yang dikeluarkan oleh Badan Perwakilan Rakyat bersama Pemerintah, serta badan Tata Usaha Negara, baik ditingkat pusat maupun daerah yang mengikat secara umum.

2.3 Definisi Data Mining

Ada banyak definisi *data mining* yang diperoleh dari jurnal ataupun buku yang ada. Adapun definisi *data mining* antara lain sebagai berikut:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Menurut Han 2006 yang dikutip dari London: Morgan Kauffman Publishers, menyebutkan bahwa *data mining* adalah suatu proses identifikasi informasi yang didapat dari berbagai *database* besar menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning*.
2. Menurut Larose 2005, yang dikutip dari John Wiley and Sons, Inc., Hoboken, New Jersey, menyebutkan bahwa *data mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda sebelumnya.

Berdasarkan definisi yang telah disampaikan, hal terkait dengan data mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang diproses berupa data yang sangat banyak.
3. Menentukan hubungan atau pola yang mungkin memberi indikasi yang bermanfaat.

2.3.1 Tahap-Tahap Data Mining

Sebagai suatu rangkaian proses, *Data Mining* dapat dibagi menjadi beberapa tahapan, yaitu:

1. Seleksi Data (*Data Selection*)
Proses pemilihan data, dimana dalam satu *database* data yang dipakai hanya data yang sesuai untuk dianalisa.
2. Pembersihan Data (*Data Cleaning*)
Proses dimana data yang tidak relevan dibuang. Data data yang diperoleh dari *database* tertentu biasanya memiliki data yang tidak lengkap, seperti ada beberapa isian yang hilang dan ada juga beberapa data yang tidak valid serta salah ketik.
3. Transformasi Data (*Data Transformation*)
Proses dimana data diubah sesuai dengan format yang ada dalam *data mining*. Seperti merubah data dalam bentuk *numeric* menjadi kategorikal untuk metode asosiasi dan *clustering*.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. Normalisasi

Pada proses normalisasi data atribut dibuat dalam skala tertentu agar menjadi kisaran data yang lebih kecil dengan tujuan agar sebaran datanya tidak terlalu jauh. Pada penelitian setiap atribut yang memiliki nilai akan di ubah menjadi 0 dan 1.

Adapun rumus untuk perhitungan normalisasi adalah sebagai berikut:

- a. *Min-Max* adalah melakukan transformasi linear terhadap data asli.

Rumus min-max adalah

$$\frac{X_{new} - (X - X_{min}) * (X_{newmax} - X_{newmin})}{(X_{max} - X_{min}) + X_{newmin}} \dots \dots \dots (2.1)$$

- b. *Z-Score* adalah normalisasi yang berdasarkan pada nilai rata-rata (*mean*) dan devisi standar (*standard deviation*) dari suatu data. Metode ini berguna jika kita tidak mengetahui nilai actual minimum dan maksimum data. Rumus dari *Z-Score* adalah

$$\frac{X_{new} - (X_i - X_{mean})}{std} \dots \dots \dots (2.2)$$

- c. *Decimal Scalling* adalah metode yang melakukan normalisasi dengan menggerakkan nilai decimal dari data ke arah yang di inginkan. Rumus dari *Decimal Scalling* adalah

$$\frac{X_{new} - X_i}{10^i} \dots \dots \dots (2.3)$$

- d. *Sigmoid Biner* adalah fungsi yang memiliki sifat non-linear. Data ditransformasi pada interval (0.1 - 0.9) karena fungsi ini merupakan fungsi asismotik yang nilainya tidak pernah mencapai 0 atau 1. Rumus *Sigmoid biner* adalah

$$y = f(x) = \frac{1}{1 + e^{-x}} \dots \dots \dots (2.4)$$

- e. *Discretization* adalah nilai nilai baku atribut numerik diubah menjadi data dengan interval label.

- f. *Concept Hierarchy Generation For Nominal Data* adalah mendefinisikan banyak hirarki untuk atribut normal yang di implisit dalam skema *database* secara otomatis.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

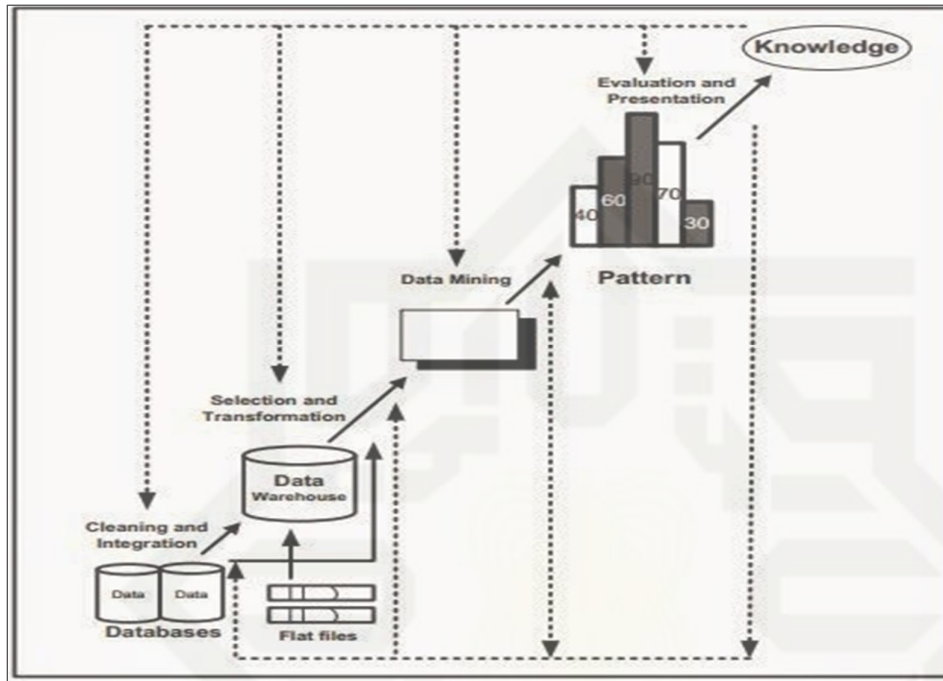
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5. Proses Mining

Proses utama ketika suatu metode diterapkan sehingga memperoleh pengetahuan baru dari data yang ada.



Gambar 2. 1 Tahapan dalam proses data mining (Han, 2006 yang dikutip dari London: Morgan Kauffman Publishers)

2.3.2 Metode dalam Data Mining

Secara umum ada beberapa metode yang terdapat didalam *data mining*, diantaranya adalah sebagai berikut:

1. Deskripsi

Deskripsi merupakan metode yang digunakan untuk menyajikan pola dan trend yang ada dalam data.

2. Estimasi

Metode ini didasarkan pada kelengkapan data (*target, variable* dan *predictor variable*). Target variable adalah variable yang diestimasi. Sedangkan predictor variable adalah variable yang dijadikan sebagai pertimbangan untuk mengestimasi target variable. Pada metode estimasi target variable harus berupa data numerik.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Klasifikasi

Metode ini bertujuan untuk membagi data ke dalam kelas-kelas tertentu berdasarkan data lama. Ada 2 tipe variable yaitu target variable dan predictor variable. Target variable berbentuk kategorikan yang akan menjadi kelas dalam klasifikasi dan predictor variable akan menjadi dasar untuk mengklasifikasikan data ke dalam kelas yang ada.

4. Kluster (*clustering*)

Merupakan metode yang mengelompokkan objek atau atribut yang sama kedalam suatu kelas. Kluster memiliki perbedaan dengan klasifikasi, dimana kluster tidak memiliki *target variable*. Metode kluster ini bertujuan untuk menghimpun data yang memiliki kesamaan ke dalam suatu kelas dan data yang berbeda dihimpun ke dalam kelas yang lain.

5. Asosiasi

Merupakan metode yang bertujuan untuk menemukan aturan-aturan yang berkaitan dengan dua variable atau lebih.

6. Prediksi

Merupakan metode yang memiliki jangka waktu, prediksi bertujuan untuk memperkirakan suatu kejadian dimasa yang akan datang.

2.4 *Support Vector Machine*

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vladimir Vapnik, Bernhard Boser dan Isabelle Guyon pada tahun 1992 di Annual Workshop on Computational Learning Theory. SVM merupakan 10 kategori metode terbaik dalam data mining, metode ini menjadikan SVM sebagai metode baru yang menjanjikan untuk klasifikasi data, baik linear maupun non-linear. Konsep dari SVM merupakan kombinasi dari konsep komputasi yang sudah ada puluhan tahun sebelumnya, seperti hyperplane (Duda dan Hart, 1973, cover 1965, Vapnik, 1964). Kernel di perkenalkan oleh Aronszajn tahun 1950.

Hak Cipta Dilindungi Undang-Undang

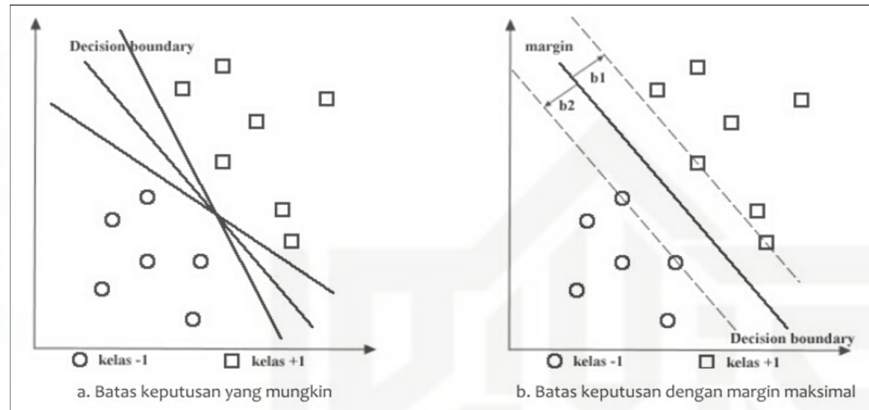
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

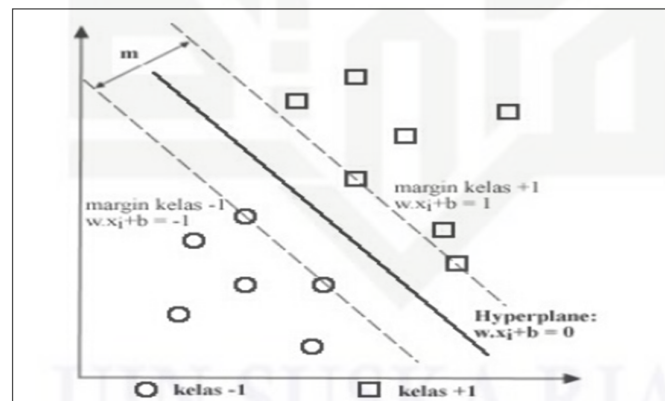
2.4.1 Konsep SVM

Konsep dasar SVM adalah memaksimalkan batas hyperplane (garis pemisah). Seperti pada gambar 2.2 berikut ini:



Gambar 2.2 Batas keputusan yang mungkin untuk set data

Pada gambar (a) ada sejumlah pilihan hyperplane yang mungkin untuk set data dan gambar (b) merupakan hyperplane dengan margin yang paling maksimal. Usaha untuk mencari lokasi hyperplane merupakan inti dari proses pelatihan SVM.



Gambar 2.3 Margin hyperplane

2.4.2 SVM Linear

Setiap data latih dinyatakan oleh $x_i.y_j$ dengan $I = 1, 2, \dots, N$, dan $x_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}^T$ merupakan atribut set untuk data latih ke-i. untuk $y_i \in \{-1, +1\}$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

menyatakan label kelas. Hyperplane klasifikasi linear SVM memiliki rumus sebagai berikut.

$$w \cdot x_i + b = 0 \dots\dots\dots(2.5)$$

Dimana w dan b adalah parameter model dan x_i adalah inner-product antara w dan b .

Data x_i yang masuk ke dalam kelas -1 adalah data yang memenuhi pertidaksamaan berikut ini

$$w \cdot x_i + b \leq -1 \dots\dots\dots(2.6)$$

Sementara data x_i yang masuk ke dalam kelas +1 adalah data yang memenuhi pertidaksamaan berikut ini

$$w \cdot x_i + b \geq +1 \dots\dots\dots(2.7)$$

Sesuai dengan gambar 2.3, jika ada data dalam kelas -1 misal (x_a) bertempat di hyperplane akan memenuhi persamaan 2.2. Untuk data kelas -1 dirumuskan

$$w \cdot x_a + b = 0 \dots\dots\dots(2.8)$$

Sementara kelas +1 dirumuskan menjadi

$$w \cdot x_b + b = 0 \dots\dots\dots(2.9)$$

Dengan mengurangkan persamaan (2.9) dengan (2.8) maka didapatkan:

$$w \cdot (x_b - x_a) = 0 \dots\dots\dots(2.10)$$

$x_b - x_a$ adalah vektor paralel di posisi hyperplane dan diarahkan dari x_a ke x_b .

Karena inner-product bernilai nol, arah w harus tegak lurus terhadap hyperplane (Gambar 2.3). dengan memberi label -1 untuk kelas pertama dan +1 untuk kelas kedua, maka untuk prediksi semua data uji menggunakan rumus:

$$y = \begin{cases} +1 & \text{jika } w \cdot z + b > 0 \\ -1 & \text{jika } w \cdot z + b < 0 \end{cases} \dots\dots\dots(2.11)$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Sesuai dengan gambar 2.3 hyperplane untuk kelas -1 (garis putus-putus) adalah data pada support vector yang memenuhi persamaan

$$w \cdot x_a + b = -1 \dots\dots\dots(2.12)$$

Sementara hyperplane kelas +1 (garis putus-putus) adalah data yang memenuhi persamaan:

$$w \cdot x_b + b = +1 \dots\dots\dots(2.13)$$

Dengan demikian maka margin dapat dihitung dengan mengurangkan persamaan (2.13) dengan (2.12), dan didapatkan:

$$w \cdot (x_b - x_a) = 2 \dots\dots\dots(2.14)$$

Margin hyperplane diberikan pada jarak antara dua hyperplane dari dua kelas tersebut. Rumus diatas diubah menjadi

$$\|w\| \times d = 2 \text{ atau } d = \frac{2}{\|w\|} \dots\dots\dots(2.15)$$

2.4.3 Hyperplane SVM

Klasifikasi kelas data pada SVM pada persamaan (2.6) dan (2.7) dapat digabungkan dengan rumus:

$$y_i (w \cdot x_i + b) \geq 1, i = 1, 2, \dots, N \dots\dots\dots(2.16)$$

Margin optimal dihitung dengan memaksimalkan jarak antara hyperplane dan data terdekat. Jarak ini dirumuskan dengan persamaan (2.15) ($\|w\|$ adalah vector bobot w). selanjutnya masalah ini dirumuskan ke dalam problem Quadratic Programming (QP) dengan meminimalkan invers persamaan (2.15), $\frac{1}{2} \|w\|^2$, dibawah konstrain seperti berikut:

Minimalkan:

$$\frac{1}{2} \|w\|^2 \dots\dots\dots(2.17)$$

Syarat:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, N \dots \dots \dots (2.18)$$

Optimalisasi ini dapat diselesaikan dengan Lagrange multiplier:

$$Lp = \frac{1}{2} ||\mathbf{w}'||^2 - \sum_{i=1}^N \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \dots \dots \dots (2.19)$$

α_i adalah Lagrange multiplier yang berhubungan dengan x_i . Nilai α_i adalah nol atau positif. Untuk meminimalkan Lagrangian, maka persamaan (2.19) harus diturunkan terhadap w dan b dan diset dengan nilai nol untuk syarat optimai di atas.

Syarat 1:

$$\frac{\partial Lp}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i y_i x_i \dots \dots \dots (2.20)$$

Syarat 2:

$$\frac{\partial Lp}{\partial w} = 0 \rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \dots \dots \dots (2.21)$$

N adalah jumlah data yang menjadi support vector.

Karena lagrange multiplier (α) tidak diketahui nilainya, maka persamaan diatas tidak diselesaikan secara langsung untuk mendapatkan nilai w dan b . Langkah yang akan dilakukan adalah dengan memodifikasi persamaan (2.19) menjadi kasus memaksimalkan dengan syarat optimal untuk dualitasnya menggunakan konstrain Karush-Kuhn-Tucker (KKT) sebagai berikut:

Syarat 1:

$$\alpha_i [y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1] = 0 \dots \dots \dots (2.22)$$

Syarat 2:

$$\alpha_i \geq 0, i = 1, 2, \dots, N \dots \dots \dots (2.23)$$

Konstrain di atas menyatakan bahwa lagrange multiplier α_i harus nol kecuali data latih x_i yang memenuhi persamaan:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) = 1 \dots\dots\dots(2.24)$$

Selanjutnya adalah dengan menyederhanakan persamaan (2.19) yaitu dengan cara transformasi ke dalam fungsi lagrange multiplier itu sendiri (dualitas masalah)

Persamaan lagrange multiplier (2.19) dapat dijabarkan menjadi:

$$L_p = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i) - b \sum_{i=1}^N \alpha_i y_i + \sum_{i=1}^N \alpha_i \dots\dots\dots(2.25)$$

Syarat optimal pada persamaan (2.21) ada suku ketiga pada ruas kanan atas pada persamaan (2.25) dan memaksa suku ini menjadi sama dengan nol. Dengan mengganti \mathbf{w} dari syarat (2.20) dan suku $\|\mathbf{w}\|^2 = \mathbf{w}_i \cdot \mathbf{w}_j$, maka persamaan di atas akan berubah menjadi dualitas lagrange multiplier L_d dan didapatkan:

$$L_d = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j \dots\dots\dots(2.26)$$

$x_i \cdot x_j$ merupakan dot-product dua data dalam data latih.

Syarat 1:

$$\sum_{i=1}^N \alpha_i y_i = 0 \dots\dots\dots(2.27)$$

Syarat 2:

$$\alpha_i \geq 0, i = 1, 2, \dots, N \dots\dots\dots(2.28)$$

Untuk set data yang besar, masalah dualitas optimasi tersebut (2.26, 2.27, 2.28) dapat diselesaikan dengan metode *numeric* seperti Quadratic Programming. Sekali α_i didapatkan, persamaan (2.20) dan (2.21) bisa digunakan untuk mendapatkan solusi layak untuk \mathbf{w} dan b .

Hyperplane (batas keputusan) didapat dengan formula:

$$(\sum_{i=1}^N \alpha_i y_i x_i \cdot \mathbf{z}) + b = 0 \dots\dots\dots(2.29)$$

N adalah jumlah data yang menjadi *support vector*, α_i merupakan *support vector*, \mathbf{z} merupakan data uji yang akan diprediksi kelasnya, dan $x_i \cdot \mathbf{z}$ merupakan inner-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

product antara α_i dan z . Untuk nilai b didapat dari persamaan (2.22) pada support vector. Karena α_i dihitung dengan teknik metode numeric dan mempunyai error numeric, nilai yang dihitung untuk b bisa jadi tidak sama. Hal itu disebabkan support vector yang digunakan dalam persamaan (2.22) biasanya diambil nilai rata-rata dari b yang didapat untuk menjadi parameter *hyperplane*. Untuk persamaan (2.22) dalam mendapatkan b dapat disederhanakan menjadi:

$$b_1 = 1 - y_1 (w \cdot x_1) \dots\dots\dots(2.30)$$

Penjelasan diatas berdasarkan asumsi bahwa kedua kelas dapat terpisah secara sempurna oleh *hyperplane*. Akan tetapi, pada umumnya kedua kelas tersebut tidak dapat terpisah secara sempurna. Hal ini menyebabkan proses optimasi tidak dapat diselesaikan karena tidak ada w dan b yang memenuhi pertidaksamaan (2.18). Untuk itu, pertidaksamaan tersebut dimodifikasi dengan memasukkan slack variabel ξ_i ($\xi_i \geq 0$) menjadi:

$$y_i (w \cdot x_i + b) \geq 1 - \xi_i \dots\dots\dots(2.31)$$

Demikian juga untuk masalah persamaan (6.13):

$$\frac{1}{2} ||w||^2 + C \sum_{i=1}^N \xi_i \dots\dots\dots(2.32)$$

Paramer C berguna untuk mengontrol trade-off antara margin dan error klasifikasi. Semakin besar nilai C maka semakin besar pula pelanggaran yang dikenakan untuk tiap klasifikasi.

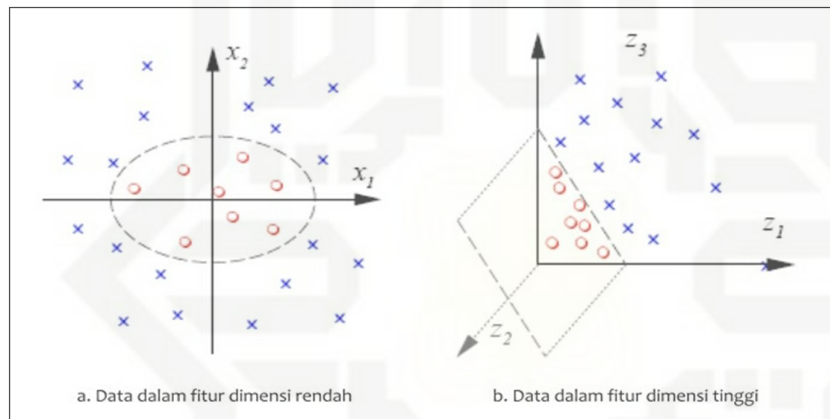
Metode untuk melakukan optimasi *hyperplane* SVM umumnya adalah untuk menyelesaikan *Quadratic Programming* dengan konstrain yang ditetapkan. Beberapa pilihan metode yang bisa digunakan adalah *chunking* (Vapnik, 1982), metode dekomposisi (Osuna *et al*, 1997), *Sequential Minimal Optimization* (SMO) (Plan, 1999).

2.4.4 SVM Nonlinear

Jika dalam artificial neural network (ANN) ada perceptron dan multilayer neural network (MLP), maka dalam SVM terdapat SVM Linear dan SVM

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Nonlinear (*kernel trick*). Seperti halnya perceptron, SVM sebenarnya adalah hyperplane linear yang hanya bekerja pada data yang dapat dipisahkan secara linear. Untuk data yang distribusi kelasnya tidak linear biasanya menggunakan pendekatan kernel pada fitur data awal set data. Kernel dapat didefinisikan sebagai suatu fungsi yang memetakan fitur data dari dimensi awal (rendah) ke fitur baru dengan dimensi yang relative lebih tinggi (bahkan jauh lebih tinggi). Pendekatan ini berbeda dengan metode klasifikasi pada umumnya yang justru mengurangi dimensi awal untuk menyederhanakan proses komputasi dan memberikan akurasi prediksi yang lebih baik.



Gambar 2.4 Dimensi Data

Ilustrasi kernel yang digunakan untuk memetakan dimensi awal (yang lebih rendah) set data ke dimensi baru (yang relative lebih tinggi) ditunjukkan oleh gambar 2.4. Algoritma pemetaan kernel diberikan oleh algoritma berikut ini:

$$\Phi : D^y \rightarrow D^q$$

$$x \rightarrow \Phi(x)$$

Φ merupakan fungsi kernel yang digunakan untuk pemetaan. D merupakan data latih, r merupakan set fitur dalam satu data yang lama, q merupakan set fitur yang baru sebagai hasil pemetaan untuk setiap data latih. Sementara x merupakan data latih, dengan $x_1, x_2, \dots, x_N \in D^y$ merupakan fitur-fitur yang akan dipetakan ke fitur berdimensi tinggi q . Jika set data yang digunakan sebagai pelatihan dengan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

algoritma adalah set data dari dimensi firut yang sama r ke dimensi baru q . Misalnya untuk sampel data N :

$$(\Phi(x_1), y_1, \Phi(x_2), y_2, \dots, \Phi(x_N), y_N) \in D^q \dots \dots \dots (2.33)$$

Pemetaan fitur lama pada set data ke fitur baru merupakan analogi dari hidden layer pad ANN, dimana jumlah neuron dalam hidden layer biasanya mempunyai jumlah lebih banyak dibanding jumlah *vector* masukan.

Selanjutnya dilakukan proses pelatihan yang sama sebagaimana paa SVM Linear. Proses pemetaan pada fase ini memerlukan perhitungan *dot-product* dua buah data pada ruang fitur baru. *Dot-product* kedua buah *vector* (x_i) dan (x_j) dinotasikan sebagai $\Phi(x_i)$, $\Phi(x_j)$. Nilai *dot-product* kedua buah *vector* ini dapat dihitung secara tidak langsung, yaitu tanpa mengetahui fungsi transformasi Φ . Teknik komputasi seperti ini disebut *kernel trick*, yaitu menghitung *dot-product* dua buah *vector* di ruang dimensi baru dengan memakai komponen kedua buah *vector* tersebut di ruang dimensi asal sebagai berikut:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \dots \dots \dots (2.34)$$

Dan untuk prediksi pada set data dengan dimensi fitur yang baru didimensikan:

$$f(\Phi(x)) = \text{sign}(w \cdot \Phi(z) + b) = \text{sign} \left\{ \sum_{i=1}^N \alpha_i y_i \cdot \Phi(x_i) \cdot \Phi(z) + b \right\} \dots \dots \dots (2.35)$$

N adalah jumlah data yang menjadi support vector, α_i adalah *support vector*, dan z adalah data uji yang akan dilakukan prediksi.

Terdapat beberapa jenis fungsi kernel *Support Vector Machine* yaitu sebagai berikut.

1. Kernel Linear

Kernel linear digunakan ketika data yang akan di klasifikasikan dapat terpisah dengan sebuah garis/hyperplane. Adapun rumus kernel linear yaitu

$$k(x,y) = x \cdot y \dots \dots \dots (2.36)$$

2. Kernel Polinomial



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kernel Polinomial digunakan ketika data hanya dapat dipisah dengan garis lekung atau sebuah bidang pada ruang dimensi tinggi. Adapun rumus kernel polynomial yaitu

$$k(x, y) = (x, y + c)^d \dots\dots\dots(2.37)$$

3. Kernel Radial Basis Function (RBF)

Kernel RBF juga digunakan ketika data hanya dapat dipisah dengan garis lekung pada ruang dimensi tinggi. Rumus kernel RBF yaitu.

$$k(x^{\rightarrow}, y^{\rightarrow}) = \exp - \frac{\|x^{\rightarrow} - y^{\rightarrow}\|^2}{2\sigma^2} \dots\dots\dots(2.38)$$

4. Kernel Sigmoid

Kernel sigmoid digunakan ketika data hanya dapat dipisah dengan garis lekung pada ruang dimensi tinggi. Rumus kernel sigmoid yaitu

$$k(x^{\rightarrow}, y^{\rightarrow}) = \tanh (k < x^{\rightarrow}, y^{\rightarrow} > + \vartheta) \dots\dots\dots(2.39)$$

2.5 Evaluasi

Evaluasi merupakan kunci yang ada pada pembuatan aplikasi berbasis *data mining*. Kinerja dari model kasifikasi dapat diukur dengan tingkat akurasi. Akurasi dari sebuah klasifikasi memberikan hasil latih dengan bentuk persentase dari kelompok data latih yang diklasifikasikan dengan benar. Perhitungannya adalah (Han dkk, 2012).

$$Akurasi = \frac{\text{Jumlah data uji benar}}{\text{Jumlah data uji}} \times 100\% \dots\dots\dots(2.40)$$

2.6 Penelitian Terkait

Tabel 2.1 Penelitian Terkait

No	Nama	Judul	Metode	Kesimpulan	Perbedaan
1.	Djati dkk, 2004.	Kajian kemampuan generalisasi <i>Support Vector Machine</i> dalam pengenalan jenis <i>Splice Sites</i> pada barisan DNA.	<i>Support Vector Machine</i> .	Tingkat akurasi sebesar 95,4%.	Sama sama menggunakan metode <i>Support Vector Machine</i> , namun kasusnya berbeda.
2.	Elly Susilowati dkk, 2012.	Implementasi metode <i>Support Vector Machine</i> untuk melakukan klasifikasi kemacetan lalu lintas pada twitter.	<i>Support Vector Machine</i> .	Tingkat akurasi mencapai 90%.	Sama sama menggunakan metode <i>Support Vector Machine</i> , namun kasusnya berbeda.
3.	Fourina Ayu Novianti dkk, 2012.	Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan <i>Support Vector Machine</i> (SVM) Berdasarkan Hasil Mamografi.	<i>Regresi Logistik dan Support Vector Machine</i> .	Regresi logistik memiliki tingkat akurasi mencapai 88,72%, sedangkan SVM mencapai akurasi sebesar 94,34%	Kasus ini menggunakan dua metode yaitu <i>Regresi Logistic</i> dan <i>Support Vector Machine</i> .
4.	Puspa dkk, 2013.	<i>Support Vector Machine</i> dilative untuk mengklasifikasikan identik	<i>Support Vector Machine</i>	Tingkat akurasi sebesar 89,8%.	Sama sama menggunakan metode <i>Support Vector</i>



Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Nama	Judul	Metode	Kesimpulan	Perbedaan
		kata perkata dan non identik kata perkata			<i>Machine</i> , namun kasusnya berbeda
5.	Khatub Mustofa dkk, 2013.	Analisis pola kemiringan tulisan tangan untuk mengidentifikasi kepribadian seseorang menggunakan <i>Support Vector Machine</i> .	<i>Support Vector Machine</i> .	Tingkat akurasi dengan data yang sama antara data pelatihan dan pengujian mencapai 92,89%. sedangkan untuk data pelatihan dan pengujian yang berbeda tingkat akurasi mencapai 92,44%.	Sama sama menggunakan metode <i>Support Vector Machine</i> , namun kasusnya berbeda.
6.	Pusphita Anna Octaviani dkk, 2014.	Penerapan metode klasifikasi <i>Support Vector Machine</i> pada data akreditasi sekolah dasar kabupaten magelang.	<i>Support Vector Machine</i> .	Tingkat akurasi mencapai 93.90%.	Sama sama menggunakan metode <i>Support Vector Machine</i> , namun kasusnya berbeda.



- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Nama	Judul	Metode	Kesimpulan	Perbedaan
7.	Aris Budianto dkk, 2015.	Deteksi nomor kendaraan dengan metode <i>Connected Component</i> dan <i>Support Vector Machine</i> .	<i>Connected Component</i> dan <i>Support Vector Machine</i> .	Tingkat akurasi mencapai 78%.	Kasus ini menggunakan dua metode yaitu <i>Connected Component</i> dan <i>Support Vector Machine</i> .
8.	Raudlatul Munawarah dkk, 2016.	Penerapan metode <i>support vector machine</i> pada diagnosa hepatitis.	<i>Support Vector Machine</i> .	Tingkat akurasi dengan menggunakan kernel linear dengan presentasi benar adalah 68-83% dan kernel RBF 70-96%.	Kasus ini menggunakan dua kernel. Yaitu kernel linear dan kernel RBF.
9.	Monalisa, Salina, 2016.	Prediksi putusan perkara pertanahan menggunakan metode <i>Naïve Bayes Classifier</i>	<i>Naïve Bayes Classifier</i> .	Tingkat akurasi mencapai 83,11%.	Kasusnya sama, namun metodenya berbeda.